HAL
open science

Polar Gaussian Processes for Predicting on Circular Domains

Espéran Padonou, O Roustant

▶ To cite this version:

Espéran Padonou, O Roustant. Polar Gaussian Processes for Predicting on Circular Domains. 2015. hal-01119942v3

HAL Id: hal-01119942
https://hal.science/hal-01119942v3

Preprint submitted on 30 Oct 2015 (v3), last revised 30 Mar 2016 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Polar Gaussian Processes for Predicting on Circular Domains

E. Padonou *
STMicroelectronics
850 Rue Jean Monnet, 38920 Crolles, France

and

O. Roustant
Mines Saint-Etienne
UMR CNRS 6158, LIMOS, F-42023 Saint-Etienne, France

October 12, 2015

Abstract

Predicting on circular domains is a central issue that can be addressed by Gaussian process (GP) regression. However, usual GP models do not take into account the geometry of the disk in their covariance structure (or kernel), which may be a drawback at least for industrial processes involving a rotation or a diffusion from the center of the disk. We introduce so-called polar GPs defined on the space of polar coordinates. Their kernels are obtained as a combination of a kernel for the radius and a kernel for the angle, based on either chordal or geodesic distances on the circle. Their efficiency is illustrated on two industrial applications. We further consider the problem of designing experiments on the disk. Two new Latin hypercube designs are obtained, by defining a valid maximin criterion for polar coordinates.

Keywords: Kriging, Polar coordinates, Design of Experiments, Kernels, Uncertainty Quantification, Disk

*The authors would like to thank the company STMicroelectronics for giving the opportunity to work on this research project, and in particular H. Duverneuil. We also thank J. Blue, from Mines Saint-Etienne for useful discussions. Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. In particular we would like to thank D. Ginsbourger and N. Durrande for their relevant remarks. Some software developments were also facilitated by two R packages produced by ReDice and Dice: DiceKriging [24], kergp [6] and DiceDesign [9]. Finally we thank A. Iyassu for his help on editing.

1 Introduction

This research aims at analyzing costly computer or physical experiments on a disk. The question was motivated by two industrial problems. The first one comes from semiconductor industry where integrated circuits are produced on disks called wafers. Several technological processes such as lithography, heating or polishing, exploit the geometry of the disk through rotations or diffusions from the center. A common issue is to reconstruct a quantity of interest over the whole disk, from few measurements at specific locations. The second problem is related to air pollution modelling for environmental impact assessment. Greenhouse gas concentrations are simulated by a computer code. Among the input variables, the couple (speed, direction) of wind characteristics can be represented on a disk, the radius of which corresponds to the maximal speed. Here also, the goal is to predict the gas concentration from some simulated experiments.

Approximation problems on the disk have been considered since the works of Zernike [25] in optics. Zernike polynomials are orthogonal with respect to the usual scalar product on the unit disk, a useful property for linear models. For such models, it is shown that optimal design of experiments are included in concentric circles [5]. More recently, a stochastic model consisting of a Gaussian process (GP), also called Kriging, has been proposed for microelectronics applications [22]. Among the existing interpolation and approximation methods, Kriging models are famous for their ability to provide both accurate prediction and uncertainty quantification, as pointed out in [12]. However their performance relies on the choice of a covariance kernel, often simply called kernel hereafter. Traditional kernels do not take into account the geometry of the disk. This may be a drawback, at least for technological or physical processes involving a diffusion from the center of the disk, or a rotation.

The main aim of the paper is to propose GP models that incorporate the geometry of the disk in their covariance kernel. For that purpose, we consider the parametrization of the unit disk in polar coordinates: $\mathcal{D} = \{(\rho\cos\theta, \rho\sin\theta), \rho \in [0,1], \theta \in \mathbb{S}\}$ where $\mathbb{S}$ represents the unit circle $\mathbb{R}/2\pi\mathbb{Z}$. The idea is to define a GP on the parametrization space $\mathcal{C} = (0,1] \times \mathbb{S}$ defined by (ρ, θ). This implies constructing a kernel on a product of the Euclidean space $(0,1]$ and of the circle $\mathbb{S}$, which can be done by algebraically combining

kernels on these two spaces with sum, product or ANOVA operations for instance. The corresponding GPs will be called here polar GPs, and the usual ones based on Cartesian coordinates, Cartesian GPs.

The construction of kernels on $\mathbb{S}$ can be achieved in several ways, and is connected to the literature of directional data (see e.g. [16, 8]) and periodic functions (see e.g. [23]). One possibility is to use so-called wrapped GP, obtained by transforming a multinormal density to a periodic one by applying an operator written as an infinite sum [15]. Here we focus on simpler approaches that provide explicit kernel expressions, either by considering restriction to $\mathbb{S}$ of a 2-dimensional GP [23], involving the chordal distance on $\mathbb{S}$, or by using the recent results of Gneiting [11], involving the geodesic distance on $\mathbb{S}$. The geodesic distance on a general manifold was recently used in the context of free-form monitoring, with so-called geodesic GPs [4]. However, the goal and the approach are quite different here, where the form is fixed (the unit disk) and the geodesic distance known analytically. Furthermore, here the geodesic is relative to the manifold $\mathbb{S}$ which is only an algebraic portion of the mapped space $\mathcal{C}$.

Second, we address the issue of defining an initial design of experiments (DoE) for circular domains. Considering the space $\mathcal{C}$ of polar coordinates is natural, but standard designs cannot be used directly due to its non-Euclidean structure. By considering a valid distance, we obtain maximin Latin hypercube designs (LHD, [18]) on $\mathcal{C}$. That class of designs is recommended when the process has a physical interpretation in polar coordinates. In order to deal with more general situations, we also propose a modified version, which still has the LHD structure with respect to ρ and θ, and is well filling the disk $\mathcal{D}$.

The paper is organized as follows. Section §2 presents the background and defines notations. Section §3 introduces so-called polar GPs. Section §4 shows the strength of the approach on two real applications, in microelectronics and environments. Section §5 is devoted to designs of experiments. Two new LHDs are introduced and compared to common designs, with respect to quality criteria. Their robustness in prediction is also investigated on a set of toy functions and models. Section §6 discusses the range of applicability of polar GPs and gives perspectives for future research.

2 Background and notations

Let $\mathcal{D}$ denote the unit disk represented either in Cartesian or polar coordinates:

$$\mathcal{D} = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\} = \{(\rho \cos\theta, \rho \sin\theta), \rho \in [0, 1], \theta \in \mathbb{S}\}$$

where $\mathbb{S} = \mathbb{R}/2\pi\mathbb{Z}$ is the unit circle. In various situations, one has to predict a variable of interest which is measured at a limited number of locations in $\mathcal{D}$. For that purpose, we will consider the framework of Gaussian Process Regression [23] also called Kriging in reference to its origins in geostatistics (see e.g. [17]). The measurement locations, also called design points, will be denoted by $X = \left(\mathbf{x^{(1)}}, \ldots, \mathbf{x^{(n)}}\right)$. In Gaussian Process Regression, the observed values at X are modelled by:

$$Y_i = \mu(\mathbf{x^{(i)}}) + Z(\mathbf{x^{(i)}}) + \eta_i \tag{1}$$

where μ is a trend function, $Z \sim GP(0, k)$ is a centered Gaussian process (GP) with covariance function – or kernel – k, and $\eta_1, \ldots, \eta_n$ are Gaussian random variables representing noise. We now briefly detail the three parts of the model.

The trend function μ is deterministic and often modeled as a linear combination of basis functions. Here, Zernike polynomials [25] are good candidates since they constitute an orthogonal basis for the usual scalar product on $\mathcal{D}$. Their shape including regular patterns are suited to describe symmetries or rotations. They were recently used for predicting on a disk [22]. The first Zernike polynomials, up to order 2, are shown in Fig. 1. The reader

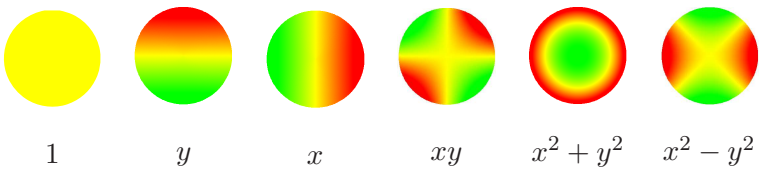


Figure 1: The six first Zernike polynomials.

is referred to [25] for more details.

The stochastic part of model 1 comprises a GP and a noise. The GP Z takes into account the spatial dependence, which thus entirely depends on its kernel k. The choice of k is crucial for applications, and may be done in order to include knowledge such as smoothness, periodicity, symmetries, etc. There are many ways to construct a kernel, and

a comprehensive presentation is found in [23], Section 4. A key idea is that multidimensional kernels can be obtained by algebraic operations, such as sum or products, of 1-dimensional kernels.

Finally the noise part represented by the η_i's may have different purposes: Modelling a measurement noise or potential discrepancies between the dataset and the kernel, and addressing numerical issues such as ill-conditioning ([1, 13]). The η_i's are computed as independent $N(0, \tau^2)$, where τ^2 is an unknown homogeneous variance term often called "nugget" or "jitter". When conditioning on the observed values, the model is an interpolator if $\tau = 0$. It is a smoother when $\tau > 0$, which gives more flexibility.

When all parameters are known, prediction with Equation (1) is given in a closed form by a Gaussian conditional distribution knowing the observations Y_i, $i = 1, \ldots, n$. Its two moments are known as Kriging mean and Kriging variance. Analytical expressions are also available when the parameters are estimated, known as universal Kriging formulas that we use here (see e.g. [23]). An important fact is that the Kriging mean at a new site $\mathbf{x}$ is obtained as an affine combination of the observed values Y_i that are correlated to $\mathrm{Z}(\mathbf{x})$. Though all the locations may be involved in the prediction, the neighboring locations, corresponding to high correlations, typically play a key role.

3 Polar Gaussian processes

One way to define a GP on the unit disk $\mathcal{D}$ is to use the restriction of a GP on the square $[0, 1]^2$, defined in Cartesian coordinates. In this paper, we will call them Cartesian GPs. In our work, we propose to further exploit the geometry of the disk by using the polar coordinates. The associated GPs will be called polar GPs.

When using the polar coordinates, the unit disk $\mathcal{D}$ is connected to the cylinder $\mathcal{C} = (0, 1] \times \mathbb{S}$, where $\mathbb{S}$ denotes the unit circle:

$$\Psi : (\rho, \theta) \in \mathcal{C} \mapsto (\rho \cos\theta, \rho \sin\theta) \in \mathcal{D} \setminus \{\mathbf{0}\} \tag{2}$$

It is a one-to-one correspondence from $\mathcal{C}$ to the unit disk without its center. The fact that the center is lost in the mapping may be a problem in theory. In practice a design point located at the center of the disk can be replaced by a set of design points placed on a closed

concentric circle. A GP on $\mathcal{D}$ can then be obtained by using Ψ^{-1}, resulting in kernels on $\mathcal{D} \times \mathcal{D}$ of the form:

$$k(\mathbf{x}, \mathbf{x}') = k_{\mathcal{C}} \left(\Psi^{-1}(\mathbf{x}), \Psi^{-1}(\mathbf{x}') \right) \tag{3}$$

where $k_{\mathcal{C}}$ is a kernel on $\mathcal{C} \times \mathcal{C}$. Such transformations are referred to as "warping" in the context of GP modeling (see e.g. [23], Section 4.2.3.).

Kernels on the cylinder can be defined by exploiting its product structure. This can be done by combining a kernel k_{r} on $(0, 1] \times (0, 1]$ and a kernel k_{a} on $\mathbb{S} \times \mathbb{S}$. A first way is by using the tensor product:

$$k_{\mathrm{prod}} (\mathbf{u}, \mathbf{u}') = k_{\mathrm{r}}\big(\rho, \rho'\big) k_{\mathrm{a}} (\theta, \theta') \tag{4}$$

where $\mathbf{u} = (\rho, \theta)$ and $\mathbf{u}' = (\rho', \theta')$ are in $\mathcal{C}$. This formulation implicitly assumes that the GP Z is the product of two independent components: a radial process R_ρ and an angular process A_θ ($Z_{\mathbf{u}} = R_\rho A_\theta$). It corresponds to a simple form of interaction. For processes that do not have interactions between these components ($Z_{\mathbf{u}} = R_\rho + A_\theta$), an additive kernel should be more appropriate:

$$k_{\mathrm{add}} (\mathbf{u}, \mathbf{u}') = k_{\mathrm{r}}\big(\rho, \rho'\big) + k_{\mathrm{a}} (\theta, \theta') \tag{5}$$

A trade-off between these two extreme approaches is the ANOVA kernel defined as:

$$k_{\mathrm{ANOVA}} (\mathbf{u}, \mathbf{u}') = \Big(1 + k_{\mathrm{r}}\big(\rho, \rho'\big)\Big)\Big(1 + k_{\mathrm{a}} (\theta, \theta')\Big) \tag{6}$$

The expanded form of Equation (6) shows that a process $Z_{\mathbf{u}}$ with ANOVA kernel can be viewed as a sum of four independent GPs: a constant process Z^0, a radial process R_ρ with kernel k_{r}, an angular process A_θ with kernel k_{a}, and a process Z^{inter} on $\mathcal{C}$ with kernel $k_{\mathrm{r}} k_{\mathrm{a}}$. From the ANOVA point of view, these processes are similar to constant term, main effects, and second-order interaction [7], but without respecting the unicity constraints such as centering. For more details on how to make new kernels from old, we refer the reader to [23].

Let us now define the kernels k_{r} on $(0, 1] \times (0, 1]$ and k_{a} on $\mathbb{S} \times \mathbb{S}$. We recall that valid kernels must be positive definite. The domain $(0, 1]$ is a segment of a 1-dimensional

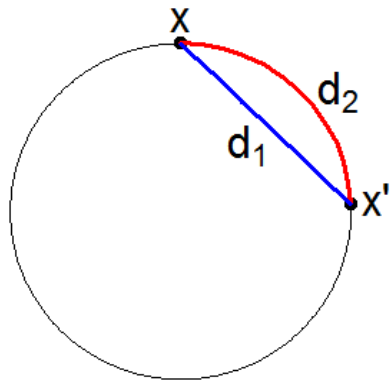


Figure 2: Chordal (d_1) and geodesic (d_2) distances on $\mathbb{S}$.

Euclidean space. As a consequence, traditional kernels are suitable for k_{r}. In particular, Matérn kernels are attractive for their ability to control the smoothness of the process and to ensure numerical stability. In dimension 1, the Matérn $\frac{5}{2}$ kernel is given by:

$$k_{\mathrm{m}}\left(x, x'\right) = \left(1 + \frac{\sqrt{5} \mid x - x' \mid}{\ell} + \frac{5(x - x')^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5} \mid x - x' \mid}{\ell}\right) \tag{7}$$

A simple way of defining kernels on $\mathbb{S}\times\mathbb{S}$ is mentioned in [11]. They are based on the chordal distance $d_1\left(\theta, \theta'\right) = 2\sin\left(\frac{\theta-\theta'}{2}\right)$ and the geodesic distance $d_2\left(\theta, \theta'\right) = \operatorname{acos}\big(\cos\left(\theta - \theta'\right)\big)$ illustrated in Figure 2.

To define a kernel on $\mathbb{S} \times \mathbb{S}$, one could be tempted to compose usual kernels with d_1 or d_2. Unfortunately, positive definiteness is not guaranteed for the resulting functions when d_2 is used. As a counter-example, if the Gaussian kernel is chosen for k_{a}, then $k_{\mathrm{a}} \circ d_2$ is not positive definite ([11], Th. 8). Alternatively, two sufficient conditions of positive definiteness over $\mathbb{S} \times \mathbb{S}$ are provided by Gneiting [11]. Define F_d the class of continuous functions $\varphi : [0, \infty) \to \mathbb{R}$, with $\varphi(0) = 1$ and such that the function $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \varphi\left(\|\mathbf{x} - \mathbf{x}'\|\right)$ is positive definite. Then:

(i). If $\varphi \in F_2$, then $\varphi \circ d_1$ is a kernel on $\mathbb{S} \times \mathbb{S}$.

(ii). If $\varphi \in F_1$ is such that $\varphi(t) = 0$ for $t \geq \pi$, then $\varphi \circ d_2$ is a kernel on $\mathbb{S} \times \mathbb{S}$.

Kernels satisfying (i) were initially proposed by Yadrenko in 1983 and are often used to describe periodic functions (see e.g. [23]). They correspond to restrictions of 2-dimensional isotropic GPs on $\mathbb{R}^2$ to $\mathbb{S}$. The second result is due to Lévy in 1961. Kernels satisfying (ii)

can be constructed from compactly supported functions on $\mathbb{R}$ such as the C^2-Wendland function defined for $0 \leq t \leq \pi$:

$$W_c(t) = \left(1 + \tau \frac{t}{c}\right)\left(1 - \frac{t}{c}\right)_+^{\tau}, \qquad c \in (0, \pi]; \tau \geq 4 \tag{8}$$

For the geodesic distance, we use $c = \pi$, which is the largest possible value due to condition (ii) above. With this choice, the covariance between two angles θ, θ' is zero when $d_2(\theta, \theta') = \pi$, and strictly positive for $d_2(\theta, \theta') < \pi$. The same interpretation is possible for the chordal distance with $c = 2$, though it is not necessary to use a compactly supported function in that case. From now on, we will use the Wendland function in both cases, resulting in the two following kernels on $\mathbb{S} \times \mathbb{S}$:

$$k_{\text{chord}}(\theta, \theta') = W_2(d_1(\theta, \theta')), \tag{9}$$

$$k_{\text{geo}}(\theta, \theta') = W_\pi(d_2(\theta, \theta')), \tag{10}$$

and the corresponding GPs will be denoted polar GP (chordal) and polar GP (geodesic).

GP simulations on the unit disk

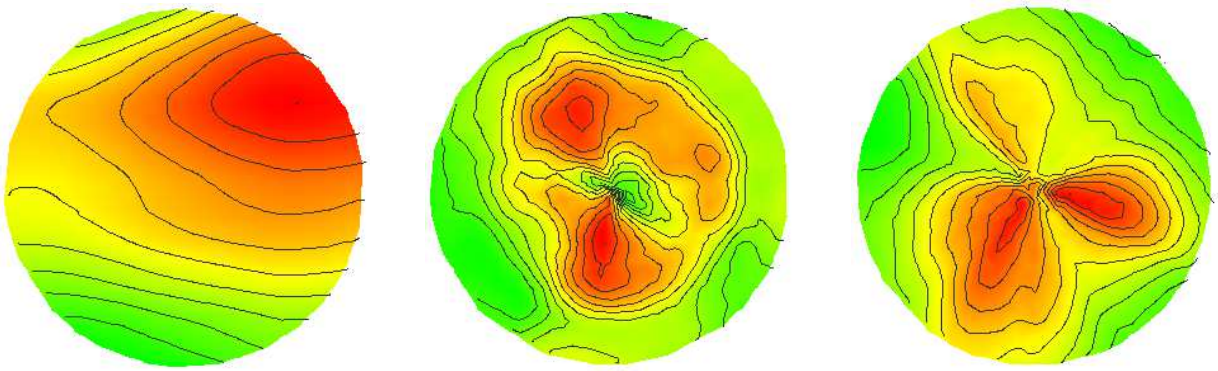
In order to better understand the specificities of polar GPs, it is useful to draw simulated surfaces. For the sake of simplicity, we propose to focus on the ANOVA combinations. We consider a Cartesian GP and the two polar GPs (chordal, geodesic) defined in Equations (9), (10). Their expressions are written below, including variance factors s^2, α_1^2, α_2^2:

(a) $k(\mathbf{x}, \mathbf{x}') = s^2\Big(1 + \alpha_1^2\, k_{\text{m}}(x, x')\Big)\Big(1 + \alpha_2^2\, k_{\text{m}}(y, y')\Big)$

(b) $k(\mathbf{x}, \mathbf{x}') = s^2\Big(1 + \alpha_1^2\, k_{\text{m}}(\rho, \rho')\Big)\Big(1 + \alpha_2^2\, k_{\text{chord}}(\theta, \theta')\Big)$

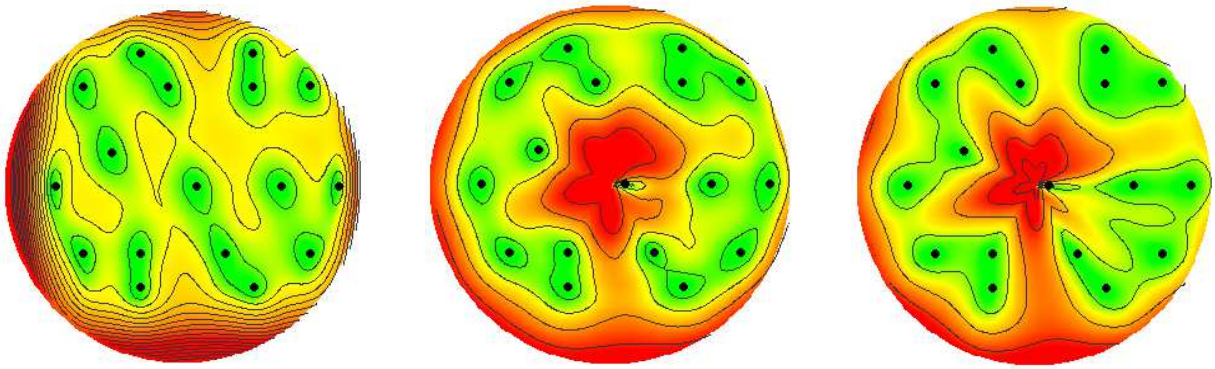
(c) $k(\mathbf{x}, \mathbf{x}') = s^2\Big(1 + \alpha_1^2\, k_{\text{m}}(\rho, \rho')\Big)\Big(1 + \alpha_2^2\, k_{\text{geo}}(\theta, \theta')\Big)$

Simulation results are displayed in Figure 3. We can see that with polar GPs, the simulated surface exhibits radial and angular patterns around the center of the disk. Such kernels may thus be suitable to describe physical phenomena involving such effects. Figure 4 shows via Kriging standard deviation how model uncertainty varies over $\mathcal{D}$, given a design of 17 points. Two striking differences are visible, especially between the Cartesian GP



Cartesian GP | Polar GP (chordal) | Polar GP (geodesic)

Figure 3: Simulations of Cartesian and polar GPs with kernels (a)-(c).



Cartesian GP | Polar GP (chordal) | Polar GP (geodesic)

Figure 4: Kriging standard deviations for Cartesian and polar GPs (kernels (a)-(c))

and the polar GP (geodesic), about uncertainty at the center of the disk, and uncertainty regions at the vicinity of design points. On one hand, the neighborhoods produced by the Cartesian GP look like elliptical regions at any location of the circular domain. On the other hand, those produced by the polar GP (geodesic) look like pie chart sectors, oriented towards the center of the disk, which plays a particular role. This is also true for the polar GP (chordal), though less pronounced.

4 Applications

4.1 Quality control in microelectronics

In microelectronics, integrated circuits are produced on circular slices of semiconductor materials called wafers. For quality monitoring, physical and electrical variables are collected on a set of locations of these wafers. In this example, the characteristic of interest is thickness, a key parameter affecting performance of integrated circuits. In our industrial background, only 17 predefined points are measured for economic reasons. The statistical challenge consists in predicting non-measured locations in order to assess the spatial risk of default from this dataset. For the purpose of this study, thickness is further measured at 64 new locations to serve as a test grid. For the sake of confidentiality, the technological process is not detailed and the thickness values are rescaled. It produces here data with a pronounced radial pattern. However, we will not assume that the model is purely radial, which is a too strong assumption in practice, due to the numerous successive operations on a wafer, and the possible slacks in processing. The aim of this section is to compare the

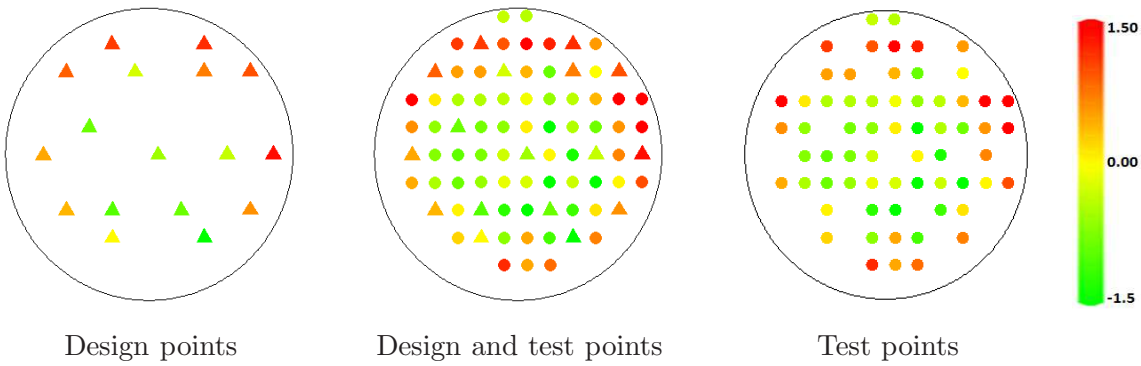


Figure 5: Rescaled thickness values. The 81 measurement locations are shown in the middle, including 17 design points (triangles, left) and 64 test points (bullets, right).

Cartesian and polar GPs (chordal, geodesic), obtained with 3 types of algebraic combination (product, sum, ANOVA). The Cartesian GPs considered here are obtained by tensor product, tensor sum or ANOVA product of the 1-dimensional Matérn kernel of Equation (7). For the polar GPs, we use the same combinations for a kernel k_{r} on $(0, 1]$ and a kernel k_{a} on $\mathbb{S} \times \mathbb{S}$, accordingly to Equations (4), (5), (6). For k_{r}, we use again the Matérn kernel,

whereas for k_{a} we choose k_{chord} or k_{geo} (see Equations (9), (10)). The range parameters τ and θ, as well as the variance factors s^2, α_1^2, α_2^2 are estimated by maximum likelihood. For this, R package kergp [6] is used with a quasi-Newton optimization (L-BFGS-B). The model accuracy is computed on the 64 test points, with the root mean squared error (RMSE) criterion. The results are summarized in Table 1 when μ is constant in Equation 1. They

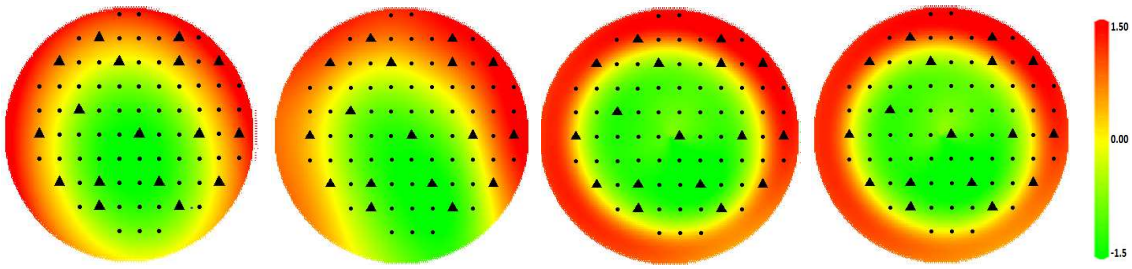
GP type	Cartesian			Polar (chordal)			Polar (geodesic)		
Kernel type	k_{prod}	k_{add}	k_{ANOVA}	k_{prod}	k_{add}	k_{ANOVA}	k_{prod}	k_{add}	k_{ANOVA}
RMSE	0.75 *	0.77	0.76	0.69	0.60 *	0.62	0.68	0.61 *	0.65

Table 1: RMSE computed on 64 test points for several GPs with a constant trend. For each GP type, the combination resulting in the smallest RMSE is marked by an asterisk. When a Zernike trend is added, the best RMSE is equal to 0.71 for all GP types, corresponding to the score of the trend only.

show that the smallest prediction errors are obtained with the polar GPs, corresponding to gains around 20% compared to the Cartesian GP. Adding Zernike polynomials as a trend slightly improves the result for the Cartesian GP, but the untrended polar GPs still outperform with a gain of 15%. Actually the trend captures the main part of the phenomenon and the GP part has then a minor effect: results are the same as for a pure linear model based on Zernike polynomials of order 2.

In order to further analyze the results, we select for each GP type the kernels corresponding to the best combination, indicated by an asterisk in Table 1. The prediction surfaces obtained with these 3 kernels are shown on Figure 6. All the GPs succeed in recovering the radial pattern of the dataset, visible on Figure 5, middle. However, it is less faithfully identified by the Cartesian GP. The differences on the predicted values can be explained by thinking at the space in which the kernel is defined. For polar GPs, prediction at one location will particularly involve the locations corresponding to a high correlation according to ρ or θ. Typically, the resulting neighborhoods in $\mathcal{D}$ may look like pie chart sectors (high radial correlation) or ring portions (high angular correlation). Here, a closer look at estimated parameters reveals that there is a high angular correlation. Therefore, prediction at the bottom of the disk involves the other points that are close to the bound-

ary. On the other hand, for the Cartesian GP, the predicted thickness has a low value, since the measurement points around, in the (x, y) space, have a low value. Finally notice that the predicted value at the extreme boundary of the disk should be considered with care, since no test points are defined on this region due to technical constraints.

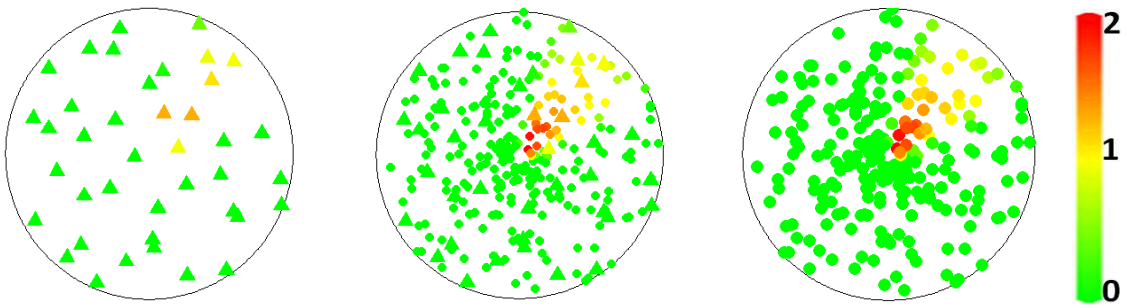


Zernike regression Cartesian GP Polar GP (chordal) Polar GP (geodesic)

Figure 6: Prediction surface for the best untrended GP models of Table 1. When adding a Zernike trend, the prediction surface is approximately the same as for a pure Zernike regression represented on the left. Black bullets correspond to test points, triangles to design points.

4.2 Air pollution modelling with a directional input

The problem tackled here is a an environmental question. A greenhouse gas emitted by a known source, usually an industrial plant, is measured at a given location for air quality monitoring. In the absence of sensors, gas concentration must be predicted. For simple landscapes, analytical expressions are available based on transport and diffusion equations. However, for complex landscapes, gas concentration is simulated by numerical codes [2]. The input variables include the emitted flow, landscape characteristics and meteorological variables. Here we focus on wind speed and wind direction. In this short study, 242 simulations were carried out, 30 of which serve as design points and the other ones are used for tests, as illustrated in Figure 7. The wind speed, initially given on the range $[0; 12]$ $(m.s^{-1})$, is rescaled to $[0, 1]$. With this transformation, the domain of the variables (speed, direction) is the unit disk. The aim of this study is simply to compare the prediction accuracy of Cartesian and polar GPs, without using a priori information. In particular, we do not specify the constraints of positivity or nullity of the gas concentration on a known



Design points. Design and test points. Test points.

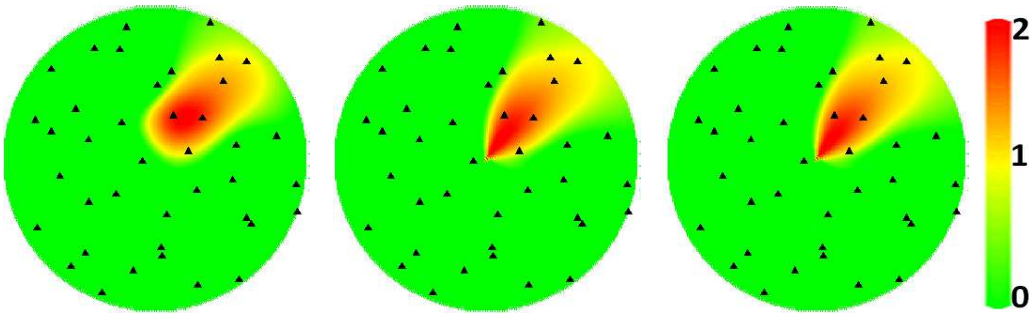
Figure 7: Rescaled gas concentrations. The 242 simulation locations are shown in the middle, including 30 design points (triangles, left) and 212 test points (bullets, right).

subregion. We use the same kernels as in the first application, corresponding to 3 algebraic combinations (product, sum, ANOVA). Here, the best model is obtained for the tensor-product combination for all kinds of GPs. This claims in favor of an interaction speed-direction for the wind on gas concentration. Notice that adding a Zernike polynomial trend does not improve the results here, since the angular shape is restricted to a region of the disk, which is hard to capture with Zernike polynomials. The results are displayed in Figure 8. In terms of prediction accuracy (measured by the RMSE criterion) the polar GPs are clearly outperforming, corresponding to gains around 40% compared to the standard tensor-product Matérn kernel. Furthermore, for the polar GPs the influence of wind direction on gas concentration has an angular shape, which is intuitive, and corresponds to the true shape visible in Figure 7 (middle). On the other hand, this shape is rectangular for the Cartesian GP.

5 Design of experiments on the disk

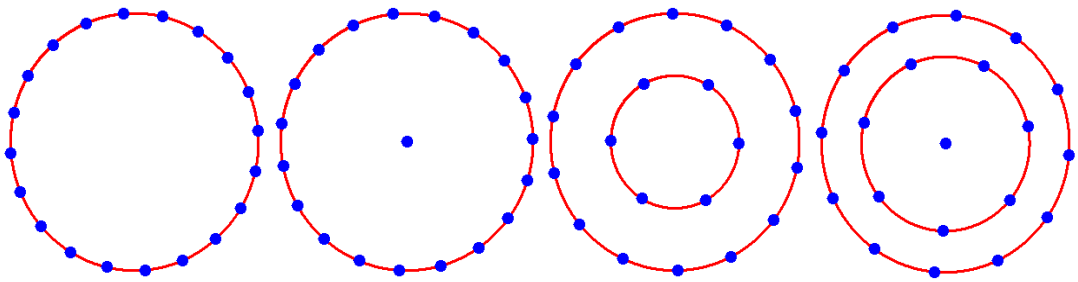
5.1 Optimal designs for Zernike polynomials and spirals

Among the DoEs that are specific to the disk, there are optimal designs for Zernike polynomials. The D-optimal designs were investigated in [5] and were found to be contained in few concentric circles, as illustrated in Figure 9.



Cartesian GP RMSE = 0.61 | Polar GP (chordal) RMSE = 0.38 | Polar GP (geodesic) RMSE = 0.37

Figure 8: Estimated gas concentrations according to wind speed (ρ) and direction (θ), for untrended Cartesian and polar GPs. Adding a Zernike polynomial trend does not improve the results. Triangles correspond to design points.

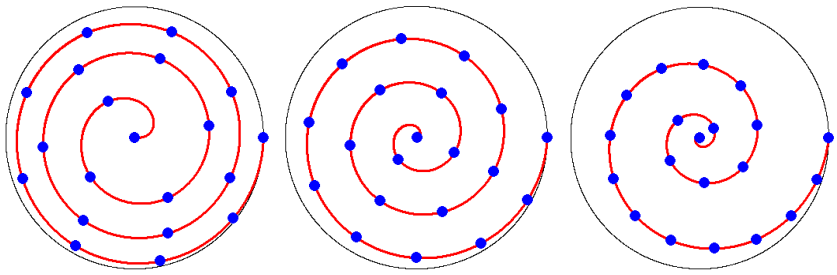


$N = 1$ $N = 2$ $N = 3$ $N = 4$

Figure 9: 20-point D-optimal DoEs for Zernike polynomials of degree N.

Spirals, hereafter denoted spiral DoEs, are used in various industrial settings: micro-electronics, optics, microbiology, etc. They allow to control the density of the design (see e.g. [21]). Some of them are represented in Figure 10, corresponding to the equation $\rho = a\theta^p + b$.

D-optimal DoEs for regression models are not robust to departures from the assumed shapes [14], and do not fill the space, a property usually required in the framework of GP modelling for capturing potential non-linearities. Poor space-filling properties are also visible for spirals in the space (ρ, θ) of polar coordinates, as shown in Figure 11, though they may correctly fill the disk.



Fermat: $s = \frac{1}{2}$ Archimede: $s = 1$ Galilee: $s = 2$

Figure 10: 20-point DoEs defined from spirals of the form $\rho = a\theta^s + b$ with $\theta \in [0, 6\pi]$. The parameter s controls the speed with which the curve moves away from the center, and a, b are chosen such that the spirals start at the center and end at the boundary.

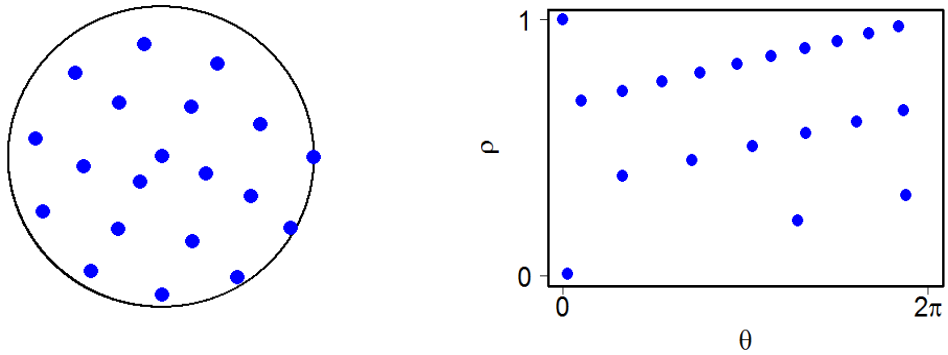


Figure 11: Cartesian (left) and polar (right) representations of the Archimedean spiral DoE. This DoE is filling well the disk but not the cylinder of polar coordinates.

5.2 Maximin Latin hypercubes for polar coordinates

For metamodelling a potentially complex phenomenon, two main properties are expected from a good DoE: Space-filling, in order to capture non-linearities, and uniformity of the marginal distributions, to avoid redundancies in projection. Among the indicators used to assess space-fillingness, the maximin criterion [19] is a common choice. In addition, Latin hypercube designs (LHD, [18]) provide good projection properties onto marginal dimensions. Thus, maximin LHDs are often proposed as initial DoEs. However such designs cannot be directly used in polar coordinates, due to the non-Euclidean structure of $\mathcal{C}$. The aim of this section is to adapt their construction.

Let us first recall the construction of a maximin LHD over the hypercubic domain

$[0,1]^2$. Given a design $\mathbf{X} = \left(\mathbf{x^{(1)}}, \ldots, \mathbf{x^{(n)}}\right)$ of elements of $[0,1]^2$, we denote $\Phi_{\text{Mn}}(\mathbf{X})$ so-called maximin criterion, giving the minimal distance among design points:

$$\Phi_{\text{Mn}}(\mathbf{X}) = \min_{i \neq j} \left(\| \mathbf{x^{(i)}} - \mathbf{x^{(j)}} \| \right) \tag{11}$$

A maximin DoE is a design that maximizes Φ_{Mn}. However, Φ_{Mn} is hard to optimize and a regularized version Φ_p, more suitable for optimization, was proposed in [20]:

$$\Phi_p(\mathbf{X}) = \left(\sum_{1 \leq i < j \leq n} \| \mathbf{x^{(i)}} - \mathbf{x^{(j)}} \|^{-p} \right)^{\frac{1}{p}} \tag{12}$$

For $p \to \infty$, maximizing Φ_{Mn} is equivalent to minimizing Φ_p. Following [20, 3], we will use $p = 50$. In software, the algorithms used for optimization are often based on simulated annealing or evolutionary strategies (see e.g. [9]). When the input variables are not provided in the same unit of measure, a maximin LHD is first designed over $[0,1]^2$, corresponding to dimensionless variables.

Now let us consider the cylinder $\mathcal{C}$ of polar coordinates. The construction of a Latin hypercube on $\mathcal{C}$ is identical for an hypercubic domain, by considering discretizations of $[0,1]$ and $\mathbb{S}$. For the sake of clarity, we propose to call *polar Latin cylinder design (polar LCD)* or simply LCD, a LHD defined in polar coordinates, referring to the geometry of the polar space. As for the maximin criterion, two modifications are needed for polar coordinates. First, a valid distance on $\mathcal{C}$ must fill the condition $\| \mathbf{u} - \mathbf{u}' \| = 0$ for $\mathbf{u} = (\rho, \theta)$ and $\mathbf{u}' = (\rho, \theta')$, with $\theta = \theta' \pmod{2\pi}$. In particular the Euclidean distance is no further valid since it does not see that the points $(\rho, 0)$ and $(\rho, 2\pi)$ are the same in $\mathcal{C}$. Second, the range of the polar angle θ is π, which is the maximum value of the geodesic distance over $\mathbb{S}$. Therefore, any distance over the dimensionless cylinder $[0,1] \times \left(\frac{1}{\pi}\mathbb{S}\right)$ applies to the polar space (ρ, θ). A natural choice is the geodesic distance given by:

$$\| \mathbf{x} - \mathbf{x}' \|_{\text{Polar}} = \sqrt{(\rho - \rho')^2 + \left(\frac{d_2(\theta, \theta')}{\pi} \right)^2} \tag{13}$$

Notice that the factor $\frac{1}{\pi}$ rescales d_2 to $[0,1]$ and weighs equivalently the radius and the angle.

From now on we will denote Φ_{Polar} (resp. $\Phi_{\text{Cartesian}}$) the Φ_p criteria computed with $\| \cdot \|_{\text{Polar}}$ (resp. $\| \cdot \|_2$). Minimizing Φ_{Polar} leads to a maximin LCD. A 20-point maximin

LCD is displayed in Figure 12, where the cylinder is represented as a 2-dimensional map. As expected it is well filling the space of polar coordinates. Though it looks similar to a

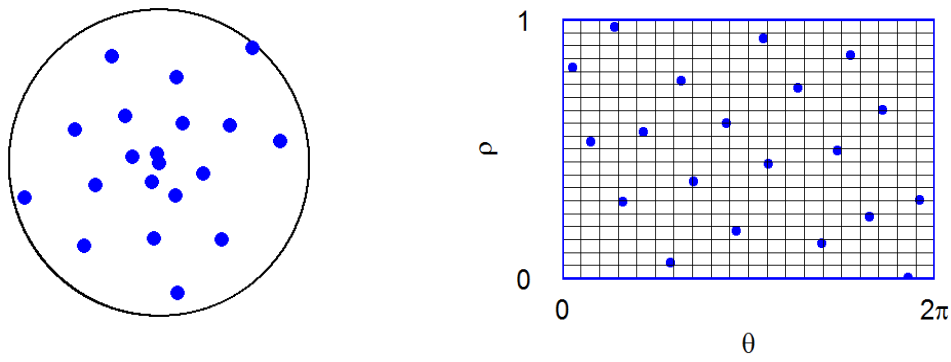


Figure 12: Cartesian (left) and polar (right) representations of a 20-point maximin Latin cylinder design (LCD). The design is well-filling the cylinder $\mathcal{C}$ of polar coordinates, displayed as a 2-dimensional map: In particular, the design points near the left and right boundaries are also spread out from each other.

maximin LHD obtained in an hypercubic domain with the usual Euclidean distance, the difference is visible on the left and right boundaries which correspond to the same points in $\mathcal{C}$: the design points near the left and right boundaries are also spread out from each other. LCDs are recommended when the studied phenomenon has a physical interpretation with respect to polar coordinates. First, if the phenomenon is purely radial (resp. angular), the Latin structure ensures that all the design radius (resp. angles) values are different, so that no information is lost by projection. Furthermore, the maximin property helps in capturing non-linearities with respect to ρ and θ. However, when no a priori information about the phenomenon is known, the maximin LCD may be inappropriate, due to non-uniform filling that they produce on $\mathcal{D}$, as visible in Figure 12. Though it is not possible to optimize simultaneously maximin criteria based on distances in Cartesian and polar coordinates, a multi-criteria approach could be investigated. In this paper, as a first study, we focus on a simple transformation of a maximin LCD which helps improving space-fillingness on $\mathcal{D}$ while preserving the Latin structure on $\mathcal{C}$. This is done by applying the transform $\rho \mapsto \sqrt{\rho}$, based on the well-known fact that if R, T are independent random variables with uniform distribution on $[0, 1]$ and $[0, 2\pi]$ respectively, then $(\sqrt{R}, T)$ is uniform on $\mathcal{D}$. This transfor-

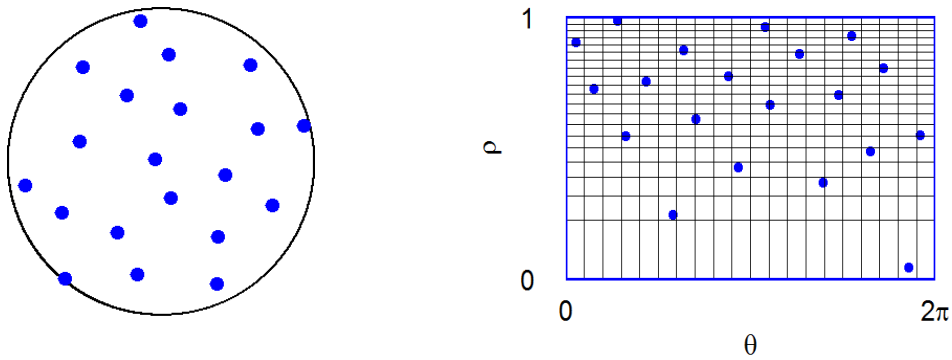


Figure 13: Cartesian (left) and polar (right) representations of the LCD obtained by transforming the maximin LCD of Figure 12 with $\rho \mapsto \sqrt{\rho}$.

mation was applied to the design of Figure 12, resulting in the design displayed in Figure 13.

5.3 Comparison

The aim of this section is to compare the DoEs presented above with respect to quality criteria, and to evaluate their performance on a set of toy functions. We will denote Dopt1, ..., Dopt4 the D-optimal DoEs for Zernike regression of order N ($1 \leq N \leq 4$) shown in Figure 9, and Spiral-F, Spiral-A, Spiral-G the spiral DoEs (Fermat, Archimede, Galilee) of Figure 10. We also denote maxLCD the maximin LCD of Figure 12 and maxLCD* its transformed version with $\rho \mapsto \sqrt{\rho}$ (Figure 13). All these 20-point DoEs are compared according to the following scheme:

(i). An assessment is made according to space-filling and D-optimality criteria. For space-filling, two indicators are used: the minimum Euclidean distance, and the minimum geodesic distance (Equation 13) between design points. The D-optimality criterion for the N-order Zernike regression (see [5]) is given in log-scale.

(ii). A comparison in term of prediction accuracy. The RMSE over a test grid of 1.000 points is computed for the 6 analytical functions shown in Figure 14, illustrating various non-linear patterns. For each DoE, the best model is chosen among Zernike polynomials up to order 4, Cartesian GPs and polar GPs with kernels obtained by

combination (sum, product, ANOVA) of 1-dimensional kernels as in Section §4.

	D-optimality			$\min_{i \neq j} \left(\| \mathbf{x}^{(\mathbf{i})} - \mathbf{x}^{(\mathbf{j})} \| \right)$	
	$D_{N=2}$	$D_{N=3}$	$D_{N=4}$	$\| \cdot \|_{\text{Polar}}$	$\| \cdot \|_2$
D-opt1	-159.9	-308.3	-448.3	0.01	0.31
D-opt2	**36.6**	-135.6	-353.1	0.01	0.33
D-opt3	35.4	**49.1**	-18.9	0.02	**0.45**
D-opt4	34.4	47.5	**63.5**	0.03	0.32
Spiral-F	29.7	37.2	44.1	0.04	0.20
Spiral-A	27.2	31.6	32.1	0.03	0.22
Spiral-G	23.3	19.3	-1.4	0.01	0.13
maxLCD	22.2	20.3	2.4	**0.06**	0.06
maxLCD*	27.5	32.8	33.0	0.04	0.28

Table 2: Comparison of DoEs according to D-optimality and space-filling criteria.

The results of Table 2 are consistent with the theory of D-optimality and exhibit the lack of robustness of D-optimal designs in case of departure from their assumptions, especially when N is underestimated. The comparison also shows that spiral DoEs have rather good scores for all criteria. The best spirals for Zernike polynomials are the one that have the smaller p (Spiral-F), but the intermediate one (Spiral-A) has the best space-filling scores; It seems to be the best trade-off among spirals. As expected, the maximin LCD is interesting for the polar GPs because it optimally fills the polar space, but has the worst space-filling score in Cartesian coordinates. This weakness is overcome by its modified version maxLCD*, which seems to accomplish the best trade-off for the different criteria among all the DoEs considered.

In Table 3, we see that D-optimal designs of low order $(1, 2)$ have in general poor scores in term of RMSE for the functions considered here, that present non-linearities. Spirals and maxLCD perform rather well. maxLCD met our expectations when radial and angular patterns are dominant (functions 1, 2, 3 and 6), and the modified maxLCD* seems to adapt well to the range of functions and models considered here, confirming its robustness among other DoEs. Finally, notice that the function $z = \sin(2\pi\rho + \theta)$ was poorly reconstructed by

	Prediction RMSE (as percentage of the standard deviation)					
	$x\left(x^2-y^2\right)$	$\left(\rho-\frac{1}{4}\right)^2$	$\sin\left(2\pi\rho+\theta\right)$	$\frac{1+\sin(\theta)}{1+\rho^2}$	$\frac{1+x}{1+y^2}$	$\cos\left(3\theta\right)$
D-opt1	14.0	210.1	112.7	75.2	2.9	2.8
D-opt2	14.4	46.6	112.4	60.2	1.7	0.2
D-opt3	0.0	17.5	80.1	18.4	**0.4**	**0.2**
D-opt4	0.0	15.9	82.1	19.0	0.8	1.4
Spiral-F	0.0	0.4	41.2	6.3	0.7	4.5
Spiral-A	0.0	0.1	49.0	4.0	0.7	2.8
Spiral-G	0.0	**0.0**	49.1	8.6	1.3	0.8
maxLCD	0.0	0.0	42.9	4.5	1.0	0.8
maxLCD*	0.0	0.3	**35.3**	**3.6**	0.4	2.6

Table 3: Comparison of DoEs in terms of predictive performance on toy functions.

all models, whatever the DoE. Indeed, either neighborhoods in (x, y) or (ρ, θ) coordinates are not suited to capture the spiral shape of the function, and an acceptable fit would require more than 20 points.

6 Conclusions and discussion

We addressed the issue of analyzing costly computer or physical experiments on a disk. Such problems are encountered in various industrial applications, where the geometry of the disk is exploited for several technological processes involving rotations or diffusions from the center. For prediction purpose, we introduced so-called polar GP models that take into account the geometry of the disk both in their mean and covariance kernel. The new kernels are defined in polar coordinates. They are obtained as a combination of a kernel for the radius using an Euclidean distance, and a kernel for the angle, based on either chordal or geodesic distances on the unit circle. It was shown in two industrial examples where radial and angular patterns are visible that the approach significantly improves prediction. The best algebraic combination was found to be either a tensor product or a tensor sum, which claims in favor of using a kernel mimicking the more general ANOVA decomposition [10]. Furthemore, in these applications there were only few differences in the results obtained with the polar GPs based on chordal or geodesic distances. This can be explained by

the strong monotonic relationship between the chordal and geodesic distance. However, in theory the geodesic distance does not distort distances on the circle, and should be preferred. Finally, though not reported here, similar results were obtained with other kernel choices such as Matérn $\frac{3}{2}$ or exponential kernels for the Cartesian GP.

We also considered the problem of designing experiments for circular domains. We introduce so-called (polar) Latin cylinder designs (LCD), which correspond to LHD in polar coordinates, and are suited to phenomena that have a physical interpretation in polar coordinates. By defining a valid distance in polar coordinates, we obtained a maximin LCD. A modified LCD was proposed for more general situations, with the additional property of filling well the disk. We assessed the performances of these two designs, compared to other common DoEs of the literature, on a set of toy functions aiming at representing various non-linear patterns. As expected, the maximin LCD is outperforming in case of radial or angular patterns, but loses efficiency otherwise. The modified LCD is attractive for its robustness, showing good performances over the whole range of functions and models considered here.

It is important to precise when polar GPs, based on distances on the unit circle, are relevant. One main difference between polar GPs and the usual ones, called here Cartesian GPs, is about the neighborhoods used for prediction. Since kernels of polar GPs are mapped to the polar space (ρ, θ), the prediction at one location will particularly involve the locations corresponding to a high radial or angular correlation with respect to ρ or θ. Typically, the resulting neighborhoods in the disk may look like pie chart sectors (high radial correlation) or ring portions (high angular correlation). This explains why polar GPs give more accurate predictions when there are radial or angular patterns, as may happen for technological processes that involve a rotation or a diffusion from the center. In other situations, involving for instance translations, Cartesian GPs may give better results. These two cases might correspond to the "two clusters of profiles over a circular grid" mentioned in [22] without any additional information about their origin. A knowledge of the process or historical data may help to choose which kernel is appropriate. In any case, there remains a lot of degrees of freedom about a GP model definition, concerning at least the trend shape or the different kernels corresponding to a given distance. To address this

problem, aggregation techniques may be a solution.

The discussion above concerning the model choice raises several questions about design of experiments. The study presented in this paper shows the possibility of adapting existing criteria to new distances. In the situations where there is no information about processing, choosing a distance may be difficult, and there is a need for building DoEs that can be suitable for any distance. In such a case, a solution would be to consider a multi-criteria approach, for instance by aggregating the maximin criteria in Cartesian and polar coordinates. On the other hand, if a specific kernel is justified, then IMSE-optimal designs could be computed with respect to this kernel.

References

[1] Ioannis Andrianakis and Peter G. Challenor. The effect of the nugget on gaussian process emulators of computer models. Comput. Stat. Data Anal., 56(12):4215–4228, December 2012.

[2] M. Batton-Hubert, M. Binois, and E. Padonou. Inverse modeling to estimate methane surface emission with optimization and reduced models: application of waste landfill plants. In 13th Annual Conference of the European Network for Business and Industrial, Ankara, Turkey, 2013.

[3] G. Damblin, M. Couplet, and B. Iooss. Numerical studies of space filling designs: optimization of Latin hypercube samples and subprojection properties. Journal of Simulation, 7(4):276–289, 2013.

[4] E. del Castillo, B. M. Colosimo, and S. D. Tajbakhsh. Geodesic Gaussian processes for the parametric reconstruction of a free-form surface. Technometrics, 57(1):87–99, 2015.

[5] H. Dette, V. B. Melas, and A. Pepelyshev. Optimal designs for statistical analysis with Zernike polynomials. Statistics, 41(6):453–470, 2007.

[6] Y. Deville, D. Ginsbourger, and N. Roustant, O. Contributors: Durrande. kergp: Gaussian Process Laboratory, 2015. R package version 0.1.0.

[7] N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro. ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. Journal of Multivariate Analysis, 115:57–67, 2013.

[8] N.I. Fisher. Statistical Analysis of Circular Data. Cambridge University Press, 1995.

[9] J. Franco, D. Dupuy, O. Roustant, G. Damblin, and B. Iooss. DiceDesign: Designs of Computer Experiments, 2014. R package version 1.6.

[10] D. Ginsbourger, O. Roustant, D. Schuhmacher, N. Durrande, and N. Lenz. On ANOVA decompositions of kernels and Gaussian random field paths. ArXiv e-prints, September 2014.

[11] T. Gneiting. Strictly and non-strictly positive definite functions on spheres. Bernoulli, 19(4):1327–1349, 09 2013.

[12] Robert B. Gramacy, Jarad Niemi, and Robin M. Weiss. Massively parallel approximate gaussian process regression. SIAM/ASA Journal on Uncertainty Quantification, 2(1):564–584, 2014.

[13] RobertB. Gramacy and HerbertK.H. Lee. Cases for the nugget in modeling computer experiments. Statistics and Computing, 22(3):713–722, 2012.

[14] P.J. Huber. Robust Statistics. Wiley Series in Probability and Statistics. Wiley-Interscience, 1981.

[15] G. Jona-Lasinio, A. Gelfand, and M. Jona-Lasinio. Spatial Analysis of Wave Direction Data using Wrapped Gaussian Processes. The Annals of Applied Statistics, 6(4):1478–1498, 2012.

[16] K. V. Mardia and P. E. Jupp. Directional Statistics. Wiley, 2000.

[17] G. Matheron. Principles of geostatistics, volume 58. Society of Economic Geologists, 1963.

[18] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2):239–245, 1979.

[19] M.D. Morris. Gaussian surrogates for computer models with time-varying inputs and outputs. Technometrics, 54(1):42–50, 2012.

[20] M.D. Morris and Mitchell T.J. Exploratory designs for computational experiments. Journal of Statistical Planning and Inference, 43:381–402, 1995.

[21] R. Navarro and J. Arines. Complete Modal Representation with Discrete Zernike Polynomials - Critical Sampling in Non Redundant Grids. INTECH Open Access Publisher, 2011.

[22] G. Pistone and G. Vicario. Kriging prediction from a circular grid: application to wafer diffusion. Applied Stochastic Models in Business and Industry, 29(4):350–361, 2013.

[23] C.E. Rasmussen and C.K.I. Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006.

[24] O. Roustant, D. Ginsbourger, and Y. Deville. DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. Journal of Statistical Software, 51(1):1–55, 2012.

[25] F. Zernike. Diffraction theory of the cut procedure and its improved form, the phase contrast method. Physica, 1:689–704, 1934.

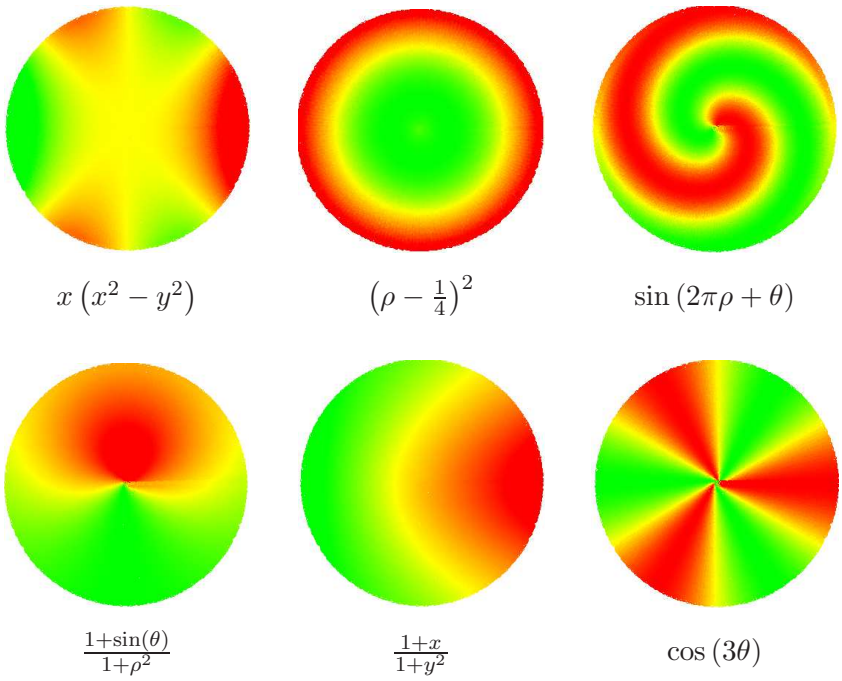


Figure 14: Color representation of test functions.